

## REPORT

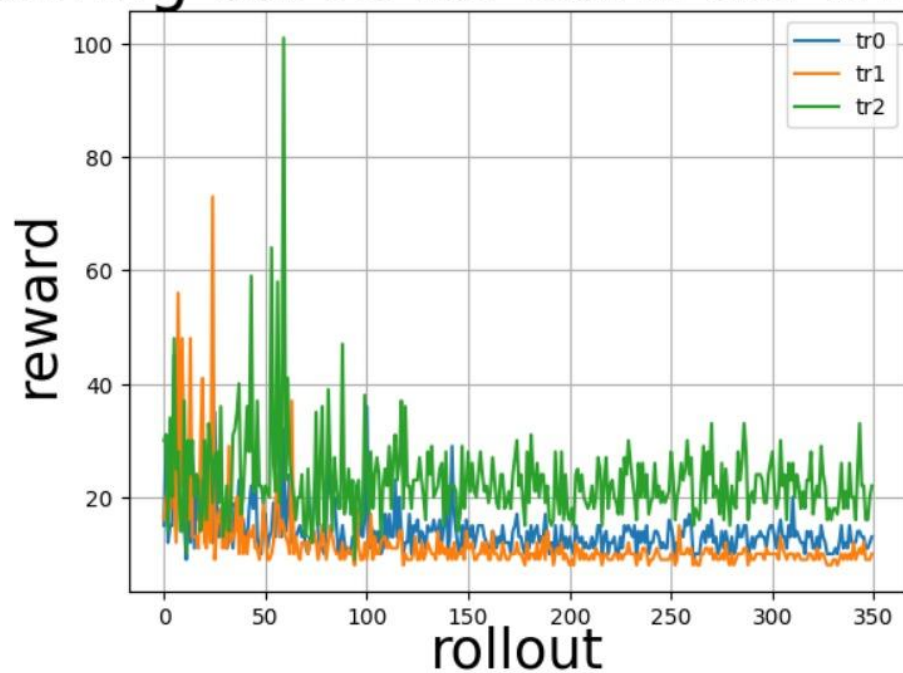
### Part 2- Experiment 1:

1. Create a graph that compares the learning curve from the four trials above. Label the curves as t0, t1, t2. Report optimum value as  $\tau_{opt}$ .

ANS:

optimal  $\tau_{opt}$  is 0.005

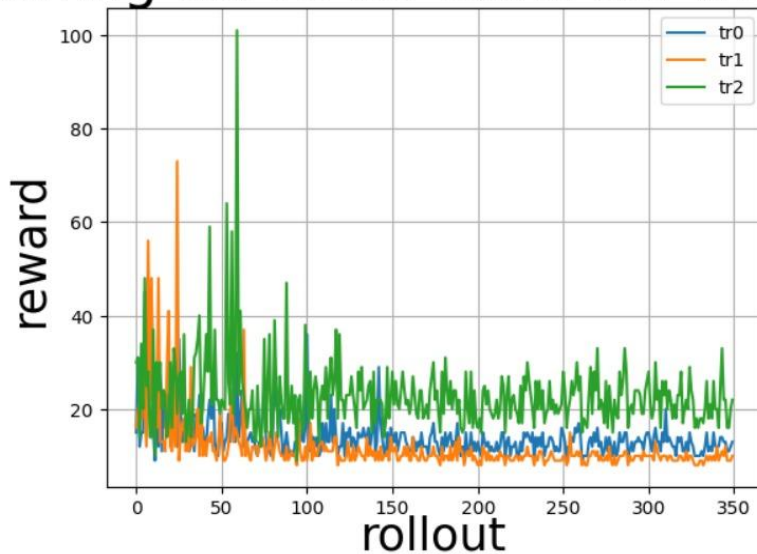
## Learning curve for CartPole with DQN



2. Create a graph that compares the learning curve from the four trials above. Label the curves as t0, t1, t2. Report optimum values as  $E_{init\ opt}$ ,  $E_{min\ opt}$ .

Ans:  $\epsilon_{initopt}$  is 0.9,  $\epsilon_{minopt}$  0.05

## Learning curve for CartPole with DQN



### 3. Answer the following questions

a) How does changing target network update rate affect the learning curve? Can you justify your observation?

b) How does changing range for  $\epsilon$  affect the learning curve? Can you justify your observation?

**Ans:** a) Changing the target network update rate can have a significant impact on the learning curve.

The target network in DQN is used to estimate the value of the next state in the learning process, and it is updated periodically to improve the stability of the learning algorithm. The update rate determines how frequently the target network is updated, which affects the rate at which the algorithm learns and the stability of the learning process.

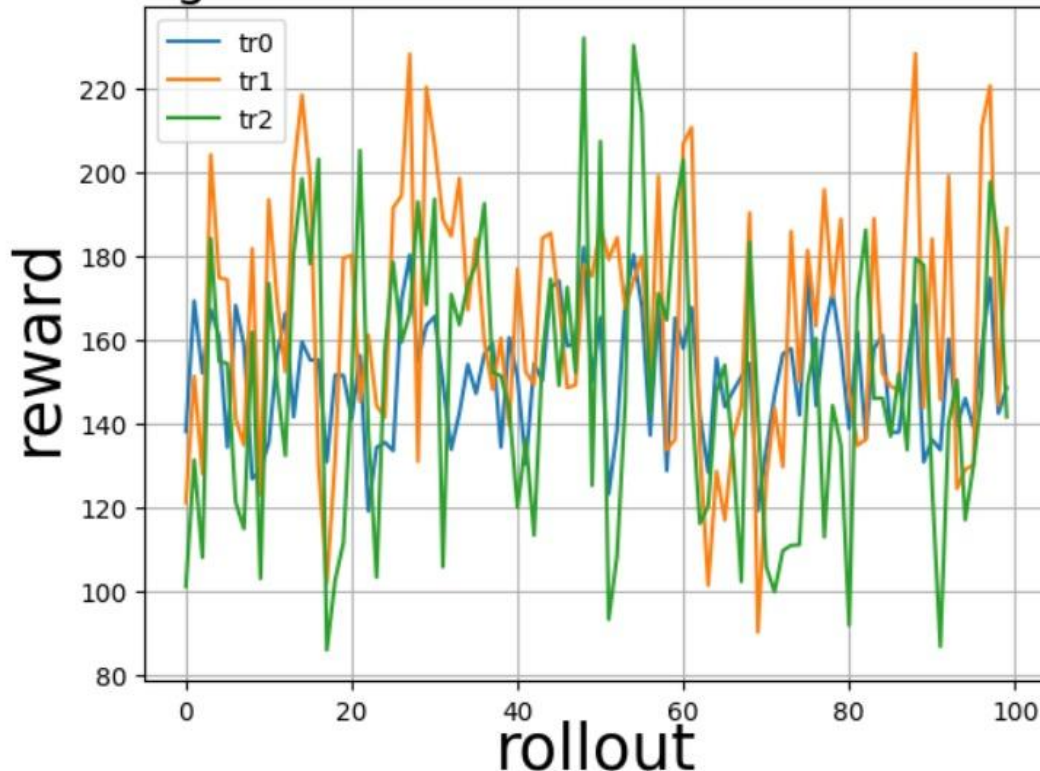
b) Epsilon( $\epsilon$ ) is the parameter that controls the exploration-exploitation trade-off during the learning process. Changing the range for  $\epsilon$ , which is typically set between 0 and 1, can have a significant impact on the learning curve.

### Part 3 - Experiment 2:

1. Create a graph that compares the learning curve from the three trials above. Label the curves as t0, t1, t2.

Ans:

Learning curve for LunarLander with Actor-Critic



2. Answer the following questions

a) How does changing the critic network update parameters (number of iterations and number of epochs) affect the learning performance? How can you justify this relationship?

Ans:

The number of iterations and epochs determines how many times the critic network is updated during each training episode, which affects the accuracy of the Q-value estimates and the stability of the learning process.

-> Increasing the number of iterations and epochs tends to make the learning process more accurate but slower. The reason for this is that the critic network becomes more

thoroughly trained with more iterations and epochs, leading to more accurate Q-value estimates. However, this increased accuracy comes at the cost of a slower learning rate, as the training process takes more time to update the critic network.

-> Decreasing the number of iterations and epochs makes the learning process faster but less accurate. With a lower number of iterations and epochs, the critic network is trained less thoroughly, which can reduce the accuracy of the Q-value estimates. This can lead to a less stable learning process, as the agent may struggle to converge to the optimal policy due to inaccurate Q-value estimates. However, the faster rate of learning can allow the agent to explore more and potentially find better solutions.

**GITHUB link:**

<https://github.com/Sreemourya710/Assignment2.git>

**Name: SreeMourya Ravirala**

**Student ID: 1002029130**